DOCUMENT RESUME

ED 050 890 RE 003 368

AUTHOR Farr, Roger
TITLE Measuring Reading Comprehension: an Historical Perspective.
PUB DATE Dec 70
NOTE 17p.; Paper presented at the National Reading Conference, St. Petersburg, Fla., Dec. 3-5, 1970
AVAILABLE FROM Twentieth Yearbook of the National Reading Conference, Inc., Marquette University, Center for Reading Services, Milwaukee, Wis. 53233 (In press)

EDRS PRICE EDRS Price MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS *Evaluation Criteria, *Historical Reviews, Measurement Instruments, *Measurement Techniques, *Reading Comprehension, Reading Processes, *Reading Tests

ABSTRACT

This review of the history of the measurement of reading comprehension follows the development of formal and informal reading comprehension tests dating from 1913 to the present. Upon reviewing the aspects, procedures, and criteria of these tests, the author noted that most of these tests resembled group verbal intelligence tests. Makeup of the tests was discussed to show how they were related to purposes for reading. It was felt that most tests had attempted to measure reading comprehension as a mere thought-getting process unrelated to reading purposes and that these tests limit the purposes for reading to the examinee's ability to achieve the test developer's purpose. A number of problems with regard to these tests were raised. It was concluded that instead of a continuous process of development, improvement, and increased knowledge in that area over the past 50 years, there had been merely a continuous rediscovery of old ideas and a continuous search for the elusive, definitive theory of reading comprehension which can serve as a basis for all measures of reading comprehension. Standardized reading tests which seem to measure reading comprehension are included in a table. Diagrams of various theories of reading comprehension and references are also included. (AW)

ED050890

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY National Reading Conference TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Measuring Reading Comprehension: an Historical Perspective

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Roger Farr*
Institute for Child Study
Indiana University

This paper is the result of an investigation of the history of the measurement of reading comprehension, undertaken to ascertain whether there has been a continuous process of development, improvement and increased knowledge in that area over the past fifty years or whether there has been merely a continuous re-discovery of old ideas and a continuous search for the elusive, definitive theory of reading which can serve as a basis for all measures of reading comprehension. As will be shown, the results of my investigation suggest that the latter more accurately describes the situation.

It would be impossible to say exactly when the measurement of reading comprehension began; for to answer that question, we would have to determine the first time a graphic symbol was interpreted by a "reader," who then demonstrated through subsequent behavior that he had understood that message. The question is obviously unanswerable.

But we can hypothesize that from the very beginning the reader's comprehension was determined in great part by his purpose for reading. Reading comprehension, before the scientific advent of measuring instruments, was probably determined by how well a reader achieved his purpose when using print as a medium.

With the scientific advent of standardized measuring instruments, the search for the "psychological construct" we refer to as reading comprehension got underway and the reader's purpose for reading was almost entirely forgotten. It is quite probable that early attempts to measure reading comprehension were based on prior development of intelligence measures. Early measures of reading comprehension

* The author wishes to thank Michael Smith for his assistance in the research for this study.

RE003 368

attempted to measure reading comprehension as if it were a general behavior that was the same under all conditions.

The very earliest formal and informal measures of reading comprehension mentioned in the published literature were based on how well a reader could reproduce what he had read. For example, in 1913 Pintner (17) reported a study which compared the oral and silent reading comprehension of fourth grade pupils. Pintner's method of measurement was to ask each child after he had read "to write down as much as he could of the matter read." In addition, Pintner's method included not allowing the examinees to look back after they had completed reading. Pintner's reproduction method is still employed today on the Silent Reading subtest of the Durrell Analysis of Reading Difficulty (8). However, on the Durrell test the students only have to recite their "memories" orally and do not have to write them.

In 1914, Brown (3) published three criteria for reading measurement: "Three things which must be accurately weighed in order to have a complete measure of reading power are: 1) rate of reading, 2) quantity of reproduction, and 3) quality of reproduction." The first reading comprehension tests published seem to meet the first and second of Brown's criteria but not the third.

It has been fairly well documented that the Gray Standardized Reading Paragraphs (12), published in 1915, was the first published reading test. However, this test included no measure of reading comprehension. Probably the first published reading comprehension measure was the Kansas Silent Reading Test (14), devised by F. J. Kelly and published about 1916.

The Kansas Reading Test resembles many of our group verbal intelligence tests of today. Two items from the list for grades 6, 7, and 8 will illustrate this point:

"A farmer puts one-half the hay from his field into the first stack, then two-thirds of what is left into a second stack. Which stack is the largest?"

"Below are two squares and a circle. If the circle is the largest of the three, put a cross in it. If one square is smaller than the circle, put a cross in the large square. If both squares are smaller than the circle, put a cross in the small square."

Reading tests today still bear a strong resemblance to group verbal intelligence tests. For example the Reading Comprehension section of the California Achievement Test (19) includes these items:

"Good morning, little boy," said the policeman. "May I help you?"

"I am lost and I cannot find my way home," said Jack. "Please help me."

14. The policeman said,
    1) "Call your father."
    2) "I am in a hurry."
    3) "I will take you home."

The sheep were playing in the woods and eating the grass. The wolf came to the woods.

15. Then the sheep
    1) went on eating.
    2) ran to the barn.
    3) ran to the wolf.

Other early reading comprehension tests, all published by 1920, were The Courtis Silent Reading Test (7), Monroe's Standardized Silent Reading Tests (16), the Haggerty Reading Examination (13), and The Chapman Reading Comprehension Test (6). The Courtis Test was a timed test in which a pupil was given three minutes to read as much as he could of a two-page story. He was then given the same story but this time it was broken into a series of short paragraphs. A set of five yes-no questions followed each paragraph and the student was given five minutes to answer as many questions as he could.

Monroe's test was a four-minute timed test in which the examinee was to read a series of paragraphs. Following each paragraph was a list of five words and the examinee was to underline the correct word according to information contained in the paragraph. An example from the test for grades 6, 7, and 8 follows:

Nanook, once so full of life, now knew perfectly well that it was all over with him. Head and tail down, the picture of resigned dejection, he stood like a petrified dog. Draw a line under the word which best describes the dog Nanook. angry - frightened - hungry - down-hearted.

A procedure for measuring rate of reading comprehension very similar to Monroe's test is still used today. The Speed and Accuracy subtest of the Gates-MacGinitie Reading Test (10) published in 1964 includes a series of paragraphs which are read by the examinee. Following each paragraph are four words, and the examinee selects the word which answers a question related to the main idea of the paragraph. For example:

Eskimos learn to turn their kayaks upside down in the water and right them again quickly. This maneuver requires courage as well as

weapons hours skill ever

The Apache Indians were nomads. They preferred following game to growing crops. They chose to fight rather than raise sheep. The Apaches were

hunters sheepherders farmers cowards

The Haggerty Reading Examination for grades 5 to 12, published about 1918, included a vocabulary test, a sentence comprehension test, and a paragraph comprehension test. The sentence comprehension test consisted of forty statements answered with a "yes" or "no." The paragraph comprehension test consisted of a series of seven paragraphs each followed by true and false statements. Timing was an important factor, as it was in all of these early tests.

The Chapman Reading Comprehension Test for grades 5 through 12, published in 1920, included a series of paragraphs. The examinee was told that the second half of each paragraph included one word which spoiled the meaning of the paragraph and a line was to be drawn through that word. For example:

"The primary characteristic of a hero is his sincerity: first and foremost he must believe in his cause. In the absence of such sincerity and without such belief, he will follow the straight path along which his hopes may be attained and his ambitions realized."

In recent years, the Chapman procedure seems to be being rediscovered in slightly different form. The Stanford High School Reading Test (9) published in 1965 and the Gates-MacGinitie Reading Test published in 1964 both utilize a procedure in which examinees are to demonstrate their comprehension of a paragraph by supplying an appropriate word to complete a sentence. In both of these tests the word is to be chosen from a given set of choices. An example from the Stanford High School Reading Test follows:

> Just as a person's family helps him and stands by him in case of need, so does his clan support him when he needs it __1__ . This __2__ may range from helping him collect the price of a bride to protecting his __3__ should he incur the wrath of other clansmen bent on blood-vengeance.

1	1 friendship 2 money	3 aid 4 strength
2	5 relationship 6 guidance	7 sharing 8 support
3	1 wife 2 possions	3 life 4 property

The examples of early reading comprehension measures above demonstrate that the first two of Brown's 1914 criteria were generally being adhered to. Rate was certainly an important aspect of each of these tests and the measures were generally attempting to determine how well an examinee could understand a written communication. However, the third criterion, quality of reproduction was neglected. By contrasting these early reading comprehension tests with our tests of today, we might wonder how far we have come in improving our measurement of the quality of reproduction. Perhaps this is because these first tests, and our tests of today, limit the purposes for reading to the examinees' ability to achieve the test developer's purpose. Indeed, reading comprehension was usually defined during the early '20's as the process of thought-getting. "This thought-getting may consist in the mere understanding of sentences or in the interpretation of paragraphs or whole selections, or it may be a combination of all these factors" (Gilliland and Jordan, 11, p. 93).

The earliest reviews of these reading comprehension measures raised many questions about what a particular test was actually measuring. For example the following criticisms of various reading comprehension tests appeared in the Buro's 1938 Mental Measurements Yearbook (4):

"A valuable feature of the tests on reading comprehension is the effort to measure the pupils' ability to make inferences from the material read. However, portions of the tests may measure intelligence rather than reading ability" (p. 131, Joseph C. Dewey -- on reviewing the Metropolitan Achievement Tests (Reading), copyright, 1931).

"The tests have been validated by customary intercorrelation with other reading tests but more especially by selecting test situations which 'represent the essential elements of the basic skills which are needed for success' in the work of the grades for which the battery is appropriate . . . . The comprehension tests require pupils (1) to follow directions, (2) to interpret meaning, and (3) to organize materials. In common with most comprehension tests, some of the items are open to criticism on the ground that they could be answered correctly by many pupils without having read the test paragraph on which they depend" (pp. 136-137, Ivan A. Brooker - on reviewing the Progressive Reading Tests, copyright 1934).

"Although the authors state that they seek to measure the ability of pupils to interpret what is read and to make inferences, it appears that the questions used for this purpose require the reproduction of facts stated directly in the reading text rather than inferences made from these facts" (p. 137, Joseph C. Dewey - on reviewing the Progressive Reading Tests, copyright 1934). "However, if the Traxler tests on comprehension are broken down, it will be shown that comprehension of the paragraphs is measured by asking for: (a) details directly stated in the content, (b) details implied in the content (involving multiple-choice technique), (c) 'yes' 'no' answers, (d) total meanings, (e) central thought, etc. These techniques have long been proven good, but this application to the paragraphs in either

of the forms is not consistent. Some paragraphs are followed by only one type of question, others by two or three, seemingly without plan. In addition, if the two forms are compared, the inconsistency is increased more than ever. Another way of putting this point is to inquire, What is paragraph comprehension? Is it something that is to be measured only through details in one instance, and a combination of several different techniques in another" (p. 139, Spencer Shank - on reviewing the Traxler Silent Reading Test, copyright 1934).

It is distressingly obvious that most of these questions are still being raised today. For example, the following quotes are from Buros Sixth Mental Measurements Yearbook (5) published in 1965:

"The first deficiency is the total lack of evidence regarding the factorial compositions of the reading tests. It is admitted that the tests measure a complex set of reading skills, but no evidence is forthcoming to support the contention that the chosen 'five major reading-for-comprehension skills' are major components of reading ability, or that the STEP reading tests do actually 'weight these five kinds of skills approximately equally.' All we know is that a committee of authorities agreed on this breakdown of reading into component skills. With due respect for the committee, it would be highly desirable to have their judgments tested and supported by empirical evidence" (p. 327, Paul Lohnes - on reviewing the Sequential Tests of Educational Progress: Reading, published in 1963).

"A useful technique is to attempt to answer reading comprehension items before reading the selection (I wish the publishers would stop calling their selections 'stories'). On Form 1, grades 7-8-9, items 59-67, this reviewer answered correctly 8 out of the 9 questions about Switzerland without looking at the passage" (p. 334, Clarence Derrick - on reviewing The Survey of Reading Achievement, published in 1959).

"The directions for administering and scoring are clear. The printing is good. Graphic profiles of subtest scores are printed on the front of each booklet. In fact, in most respects, the Developmental Reading Tests 'look' like well-made standardized tests and this is perhaps what is so insidious. The teacher follows the directions, the students mark the booklet, the tests are scored, and Johnny gets a reading grade of 1.9. What on earth does this mean?" (p. 294-295, Edward Fry-on reviewing the Developmental Reading Tests, published in 1961).

The continuous search for the elusive answer to the question of what is reading comprehension probably encouraged the development of a vast multitude of reading comprehension measures. Indeed, the period from the 1940's to today could be labelled the time of sub-skills proliferation. Many tests merely labelled the same sub-tests with different titles. Others had similar labels but employed different question types. I recently collected a list of all of the subtests from reading tests which seemed to be attempting to measure reading comprehension. Most of these tests were published during the 1950's and 60's. My list includes the fifty sub-tests listed in Table 1.

Insert Table 1 about here

What do all of the reading comprehension tests of today and those of the past seem to be measuring? In my opinion they have all attempted to measure reading comprehension as a "thought-getting process" which is generally unrelated to specific reading purposes. The series of diagrams in Table 2 illustrate the makeup of reading comprehension tests and how they are related to purposes for reading.

Insert Table 2 about here

The first diagram is an adaptation from a recent article by John Bormuth et al. (1). That diagram describes the usual procedure of having a set of questions

follow the reading of a selection. The solid lines in the diagram indicate a direct relationship while the broken line represents an implied relationship. For example, in the first diagram there is only an implied relationship between a student's response to a question and his comprehension of a reading selection. The second diagram illustrates an item in which the examinees are able to identify correct responses without reading a selection. It is amazing the vast number of questions on reading comprehension tests which exemplify this second diagram.

The third diagram indicates the Cloze procedure in which a solid line establishes the direct relationship of response to comprehension of a selection. The fourth describes the chunked test which has been developed by Carver and Darby.

Each of these first four diagrams illustrates what I believe to be the basic problem with reading comprehension measures. The tests are being developed as if there was a well known theoretical construct called reading comprehension. Some of the most recent attempts at developing reading comprehension measures have emphasized that point. Schlessinger and Weiser (19) and Bormuth (2) have been vehement advocates of more systematic approaches to the development of test items. According to these authors what is needed are item development procedures which are based on the implicit language and organizational structure of a written message. Bormuth, in fact, concludes a recent study:

"The most startling result was the fact that large proportions of the children were unable to demonstrate a comprehension of even these basic structures by which information is signaled indicating that this deficiency may constitute a serious impediment to the efficiency of instruction. The structures identified seemed to represent homogeneous classes of behavior since the variation between questions measuring different skills was significantly greater than the variation between items measuring the same skill. The fact that the structures and question types differed significantly in difficulty was also taken as evidence that many of these skills may be hierarchiacally related" (1).

The fifth diagram is the avenue which I believe may lead us off the track of trying to ascertain a general reading comprehension skill. This diagram, which is probably not a description of any existing reading test, includes a direct relation of purpose to the reading selection to a student's response to a reading selection based on the given purpose.

The approach that I am suggesting would not wait for tests to be developed until there is sound theoretical and empirical evidence concerning the components of reading ability as Kingston (15) suggests. Instead we would begin to list, and organize reading purposes and tasks that lead adults and children to printed material. Measures would be based on these purposes and tasks. For example, we might include such tasks as having a child locate a phone number or determine from a newspaper when his favorite television program is scheduled. Certainly more complex purposes and tasks such as determining a character's motives to determine why he committed a certain act should be included. However, these purposes should be developed on the basis of meaningful reading tasks; they should be related to realistic purposes and should not be tasks that are derived from some authority's arbitrary decision as to what a reader's purpose should be.

Conclusion

It cannot be denied that the sophistication of test developers and test reviewers has increased tremendously. They have definitely learned to ask more probing questions about the theoretical construct of reading comprehension; they have been able to provide more sophisticated technical data on reliability, validity and norming procedures; the editing of test items has improved dramatically. But the essential questions are still the same: we are still asking what reading comprehension is and how it should be measured. The following questions about reading comprehension tests were raised in 1910; were reiterated in 1938; and are still being asked today:

1. Why is there such a great overlap between measures of hypothesized different skills of reading comprehension?
2. What format should reading comprehension measures take: multiple choice questions? cloze? fill in? long passages or short? should examinees be allowed to look back at a selection when answering questions?
3. What are the sub-skills of reading comprehension?
4. How strong an effect does prior knowledge of a topic have on an examinee's reading comprehension?
5. Does the language structure of a selection affect reading comprehension?

I suggest that these and other similar questions result in exercises in futility. The only validity of any importance is how well a test predicts a student's ability to perform functional reading tasks. Reading measures need to be developed which are based on specific reading tasks and purposes for reading.

Just as we have been generally disillusioned with our attempts to measure intelligence as a psychological construct perhaps we should also be disillusioned with our attempts to measure reading comprehension as a psychological construct. The measurement of reading comprehension should be based on an attempt to determine how well a reader can accomplish a given task with a given reading selection. From my perspective, the history of the measurement of reading comprehension got started on a narrow, single track over fifty years ago and has been chugging around in circles ever since. That is not to say that increased sophistication in the technical, scientific, and even artistic aspects have been non-existent. Indeed, some of the advances in those aspects have been quite dramatic, but the essential problem is that the train has never switched off the initial track, it has just had a streamlined engine attached.

References

1. Bormuth, J. R. Carr, J. Manning, J. and Pearson, D. "Children's Comprehension of Between - and Within - Sentence Syntactic Structures," Journal of Educational Psychology, October 1970, 61, 350.

2. Bormuth, J. R. On the Theory of Achievement Test Items, Chicago: University of Chicago Press, 1970.

3. Brown, H. A. "The Measurement of the Efficiency of Instruction in Reading," Elementary School Teacher, 1914, 14, 477-490.

4. Buros, Oscar K., ed. The Nineteen Thirty-Eight Mental Measurement Yearbook, New Brunswick, New Jersey: Rutgers University Press, 1938.

5. Buros, Oscar K., ed. The Nineteen Sixty-Five Mental Measurement Yearbook, New Brunswick, New Jersey: Rutgers University Press, 1965.

6. Chapman, J. C. Chapman Reading Comprehension Test, Minneapolis: Educational Test Bureau, 1920.

7. Courtis, S. A. Courtis Silent Reading Test.

8. Durrell, Donald D. Durrell Analysis of Reading Difficulty, New York: Harcourt, Brace & World, Inc., 1955.

9. Gardner, E. F.; Merwin, J. C.; Callis, R.; and Madden, R. Stanford Achievement Test: High School Reading Test, New York: Harcourt, Brace & World, Inc., 1965.

10. Gates, A. I. & MacGinitie, W. H. Gates-MacGinitie Reading Test, New York, New York: Teachers College Press, 1964.

11. Gilliland, A. R. and Jordan, R. H. Educational Measurements and the Classroom Teacher, New York: The Century Co., 1925.

12. Gray, W. S. Standardized Oral Reading Paragraphs, Indianapolis: Public School Publishing Company, 1915.

13. Haggerty, M. E. & Joonan, M. E. Haggerty Reading Examination, New York: World Book Co., 1920.

14. Kelly, F. J. The Kansas Silent Reading Test, Emporia, Kansas: Bureau of Educational Measurements and Standards, 1916.

15. Kingston, A. "Is Reading What the Reading Tests Test?" in E. L. Thurston and L. E. Hafner, eds., The Philosophical and Socialogical Bases of Reading, Yearbook of the National Reading Conference, 1965, 14, 106-109.

16. Monroe, Walter S. Monroe's Standardized Silent Reading Test, Indianapolis: Bobbs-Merrill Co., Inc., 1919.

17. Pintner, Rudolph. "Oral and Silent Reading of Fourth Grade Pupils," Journal of Educational Psychology, 1913, 4, 333-337.

18. Schlesinger, I. M. and Weiser, Zehavit. "A Facet Design for Tests of Reading Comprehension," Reading Research Quarterly, Summer 1970, 4, 566-580.

19. Tiegs, E. W. & Clark, W. W. California Reading Test, Monterrey, California: California Test Bureau, 1957.

Table 1

*Subtests of Standardized Reading Tests
Which Seem to Measure Reading Comprehension

(Note: Not included are the myriad of study skills; also not included are the wide range of vocabulary subtests which might in several cases be included with this listing.)

1. Context Reading
2. Sentence and Word Meaning
3. Paragraph Meaning
4. Comprehension
5. Level of Comprehension
6. Speed of Comprehension
7. Interpretation of Reading Materials
8. Interpretation
9. Organization
10. General Comprehension
11. Specific Comprehension
12. Reading to Retain Information
13. Reading to Organize
14. Reading to Evaluate - Interpret
15. Reading to Appreciate
16. Perception of Relations
17. General Information
18. Ability to Grasp the Central Thought
19. Ability to Note Clearly Stated Details
20. Interpretation
21. Integration of Dispersed Ideas
22. Ability to Draw Inferences
23. Recalling Information
24. Reading to Locate Information
25. Reading for Description
26. Rate of Reading for Meaning
27. Reading for Directions or Procedures
28. Sentence Completion
29. Retention of Details
30. Directed Reading
31. Comprehension Accuracy
32. Reading Efficiency
33. Reading
34. Noting Details
35. Interpreting Paragraphs
36. Following Directions
37. Reading for Inferences
38. Reading for Main Ideas
39. Summarizing
40. Skimming

41. Recall of Information Read
42. Gross - Comprehension
43. Comprehension - Efficiency
44. Ability to Recall Ideas
45. Ability to Translate Ideas and Make Inferences
46. Ability to Analyze Motivation
47. Ability to Analyze Presentation
48. Ability to Criticize
49. Reading for Information
50. Story Comprehension

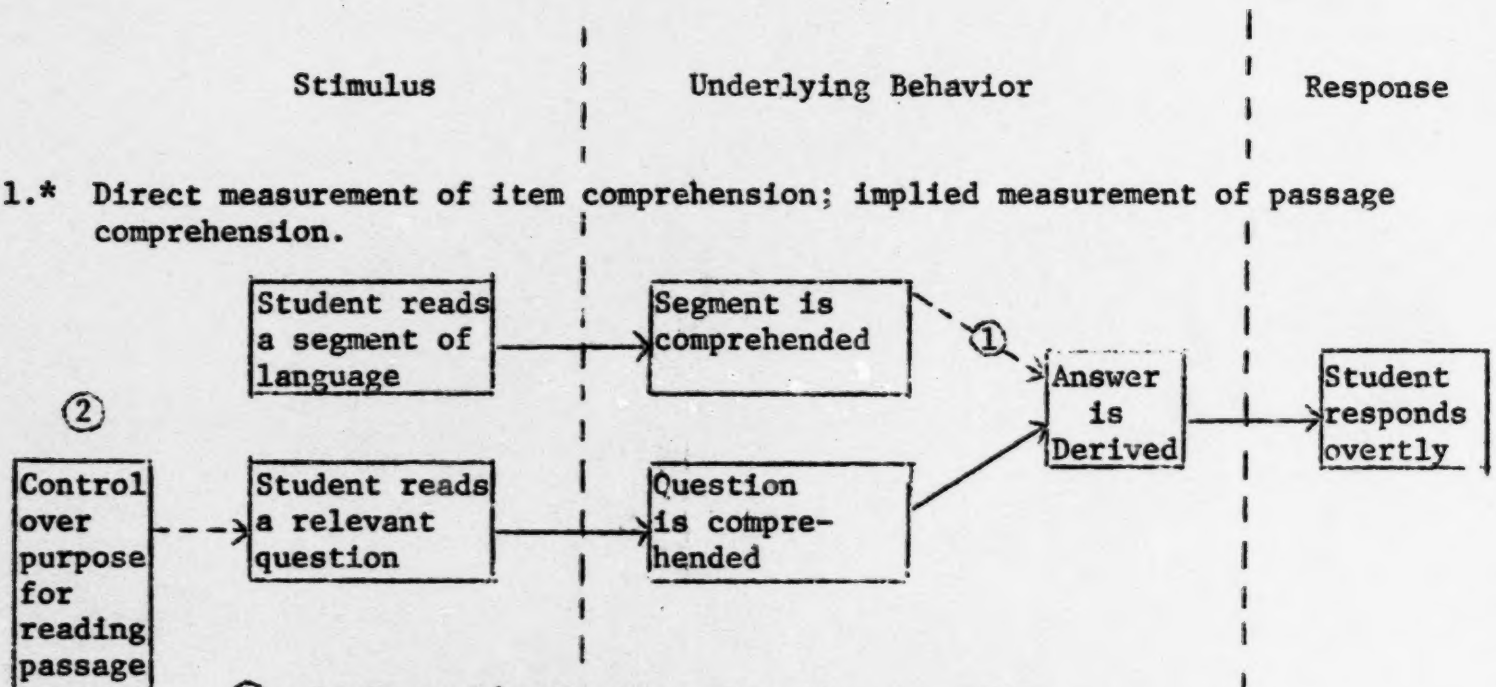
* This list was developed from the test guide appendix in:

Farr, Roger. Reading: What Can Be Measured. Newark, Delaware: International Reading Association, 1970.

Table 2

DESIGN: READING COMPREHENSION

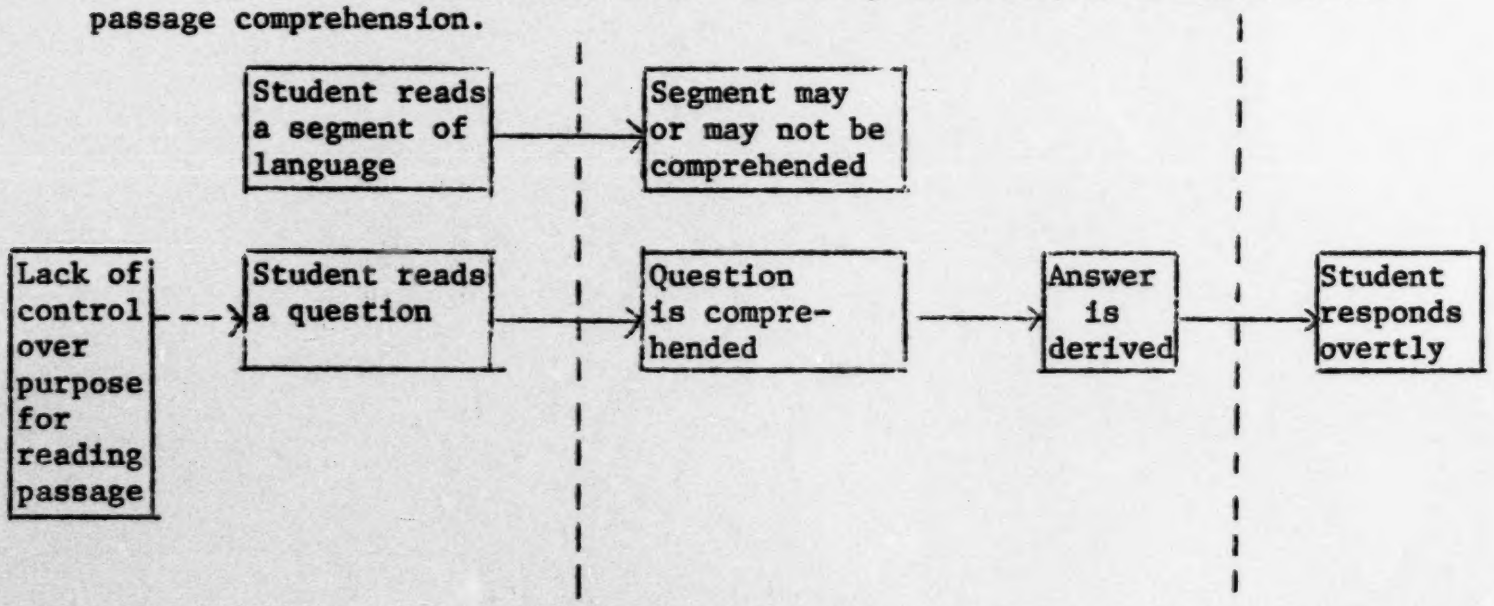
1.* Direct measurement of item comprehension; implied measurement of passage comprehension.



① In Bormuth's diagram, this line is solid.

② This block does not appear in Bormuth's diagram.

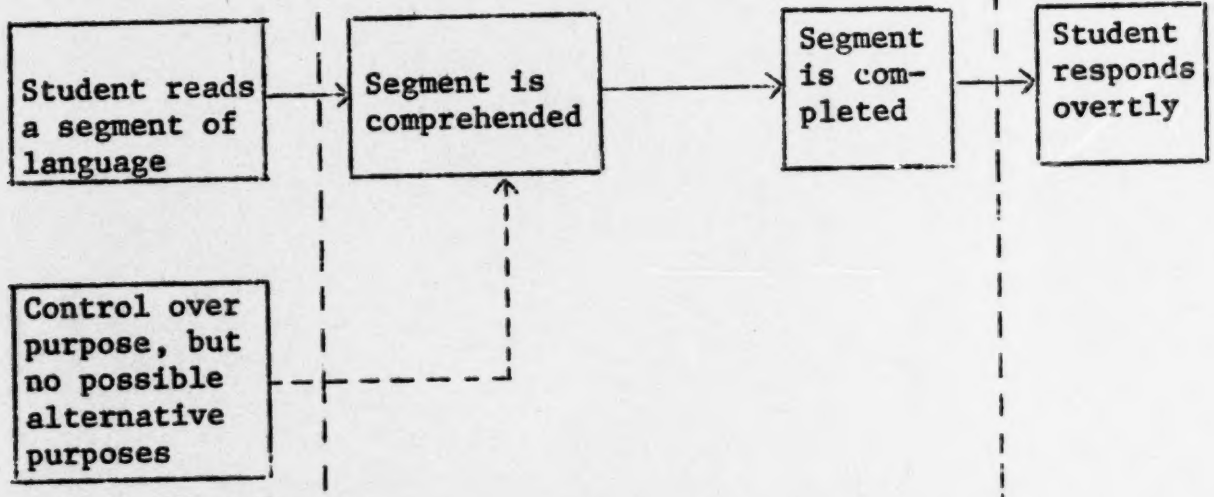
2. Poor Items: direct measurement of item comprehension; no measurement of passage comprehension.


Student reads a segment of language
Segment may or may not be comprehended
Lack of control over purpose for reading passage
Student reads a question
Question is compre-hended
Answer is derived
Student responds overtly

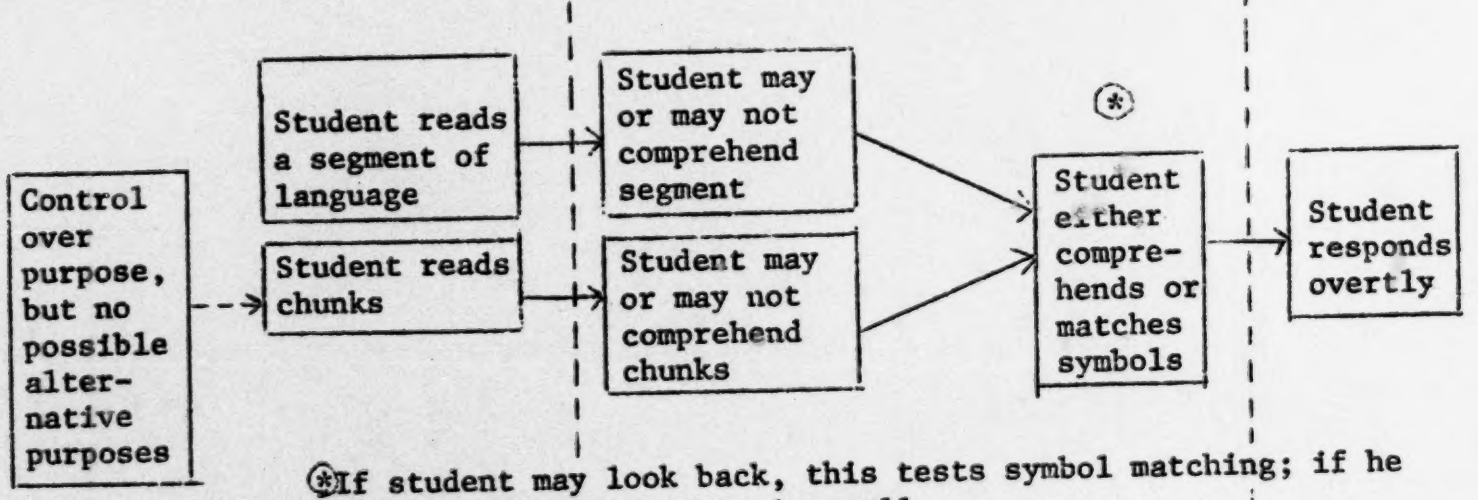
*Adapted from Bormuth, J.R., Carr, J., Manning, J. and Pearson, D., "Children's Comprehension of Between- and Within-Sentence Syntactic Structures," Journal of Educational Psychology, 61 (October 1970) p. 350.

Stimulus | Underlying Behavior | Response

3. Cloze: direct measurement of literal comprehension.

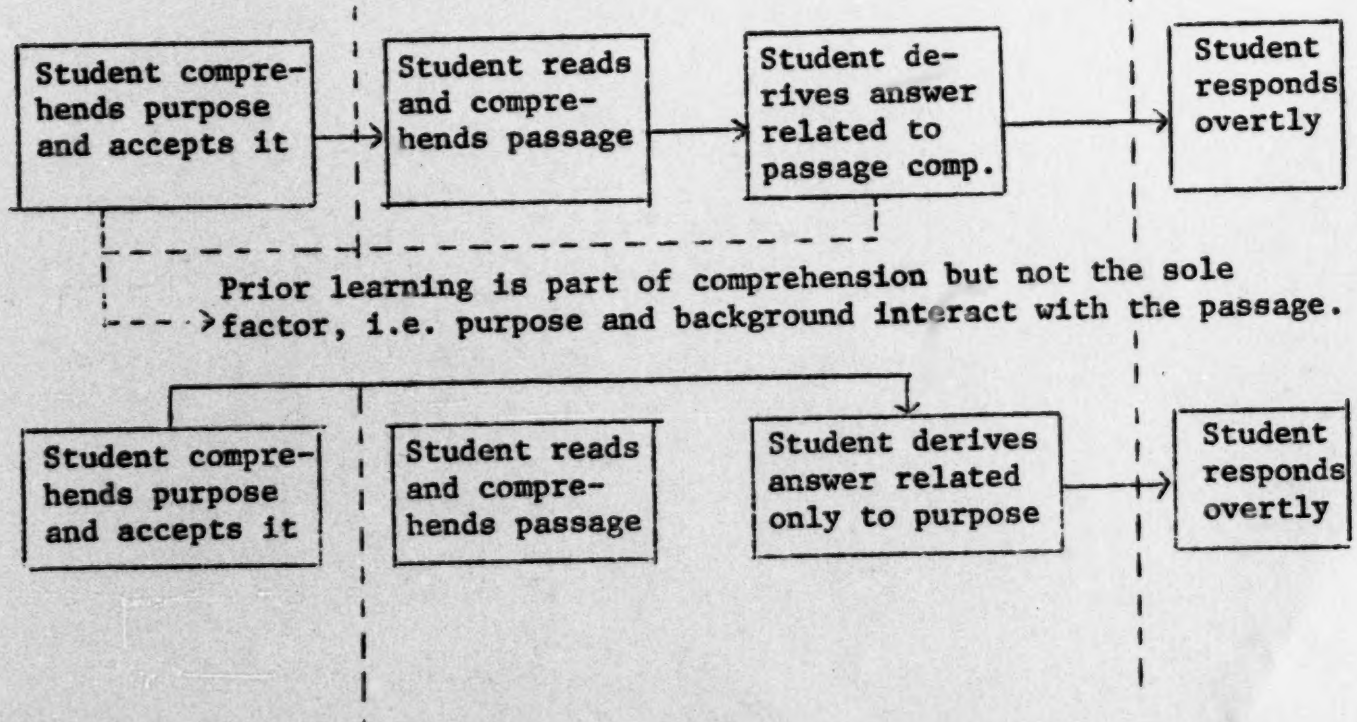

Student reads a segment of language
Segment is comprehended
Segment is completed
Student responds overtly
Control over purpose, but no possible alternative purposes

4. Chunked: implied measurement of both item and passage comprehension.


Control over purpose, but no possible alter-native purposes
Student reads a segment of language
Student reads chunks
Student may or may not comprehend segment
Student may or may not comprehend chunks
(*)
Student either compre-hends or matches symbols
Student responds overtly

(*)If student may look back, this tests symbol matching; if he may not, it tests literal recall.

5. Comprehension related to purpose.


Student compre-hends purpose and accepts it
Student reads and compre-hends passage
Student de-rives answer related to passage comp.
Student responds overtly
Prior learning is part of comprehension but not the sole factor, i.e. purpose and background interact with the passage.
Student compre-hends purpose and accepts it
Student reads and compre-hends passage
Student derives answer related only to purpose
Student responds overtly